

**TITLE: METHOD FOR GENOME MINING FOR SECRETED PROTEIN GENES**

5 **CROSS-REFERENCE TO RELATED APPLICATIONS**

This application claims, under 35 U.S.C. 119, priority of Danish application no. PA 2000 00963, filed June 21, 2000, and the benefit of U.S. application no. 60/213,376 filed June 21, 2001, the contents of which are fully incorporated herein by reference.

10

**FIELD OF THE INVENTION**

The present invention relates to a method for mining genomes for genes encoding proteins and peptides, including enzymes, that are secreted from organisms, in particular microorganisms such as bacteria and fungi. The method uses immunoassay techniques to identify clones expressing genes encoding enzymes or other proteins or peptides that are secreted from an organism of interest.

20 **BACKGROUND OF THE INVENTION**

A typical bacterial genome is about 3-8 mB (mega base pairs) in size. For example, the Bacillus subtilis genome is known to have a size of 4.2 mB and to contain a total of 4100 protein coding genes. The function of 1200 gene products of Bacillus subtilis has been experimentally identified. The function of 42% of the genes could at the time the genome sequence was finished not be predicted by similarity to known genes encoding proteins with known function. These genes could be divided into three groups: 12% showed similarity to other genes with unknown function from other organisms, while 4% showed similarity to other genes with unknown function in B. subtilis only. The remaining 26% did not show homology to

anything (F. Kunst et al. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature 320:249-256).

Screening for enzymes, or other proteins or peptides, normally involves gene cloning in order to obtain a reasonable yield of a given gene encoding a desired product. Constructing a gene library, where the genome is cut into fragments, which are then ligated into a vector and transformed into a cloning host, does this. If the genome of *B. subtilis* is randomly cut into fragments with an average size of 4 kB (kilo base pairs), at least about 1000 clones must be screened in order to cover the entire genome once. In order to ensure that all open reading frames of the genome are represented in full length, a much higher number of clones must be screened in order to ensure that the entire genome is expressed. Usually the number of screened clones is on the order of magnitude of 5000-10,000 clones.

The genomes of *Aspergillus nidulans* and *Neurospora crassa* are known to have a size of 31.0 mB and 42.9 mB of DNA, respectively (Dunn-Colemann N. & Prade, R. 1998. Toward a global filamentous fungus genome sequencing effort. Nature Biotechnology, 16, 5; Radford and Parish, 1997. The genome and genes of *Neurospora crassa*. Fungal Genetics and Biology, 21, 258-266). The nuclear genome of *Saccharomyces cerevisiae* contains 13.0 mB, and about 6200 open reading frames have been predicted (Zagulski, M., Herbert C.J. & Rytka, J 1998. Sequencing and functional analysis of the yeast genome. Acta Biochimica Polonica, 45, 627-643).

10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
100  
105  
110  
115  
120  
125  
130  
135  
140  
145  
150  
155  
160  
165  
170  
175  
180  
185  
190  
195  
200  
205  
210  
215  
220  
225  
230  
235  
240  
245  
250  
255  
260  
265  
270  
275  
280  
285  
290  
295  
300  
305  
310  
315  
320  
325  
330  
335  
340  
345  
350  
355  
360  
365  
370  
375  
380  
385  
390  
395  
400  
405  
410  
415  
420  
425  
430  
435  
440  
445  
450  
455  
460  
465  
470  
475  
480  
485  
490  
495  
500  
505  
510  
515  
520  
525  
530  
535  
540  
545  
550  
555  
560  
565  
570  
575  
580  
585  
590  
595  
600  
605  
610  
615  
620  
625  
630  
635  
640  
645  
650  
655  
660  
665  
670  
675  
680  
685  
690  
695  
700  
705  
710  
715  
720  
725  
730  
735  
740  
745  
750  
755  
760  
765  
770  
775  
780  
785  
790  
795  
800  
805  
810  
815  
820  
825  
830  
835  
840  
845  
850  
855  
860  
865  
870  
875  
880  
885  
890  
895  
900  
905  
910  
915  
920  
925  
930  
935  
940  
945  
950  
955  
960  
965  
970  
975  
980  
985  
990  
995

A screening for enzymes in fungi can be based on an expression-cloning method, which combines the ability of *Saccharomyces cerevisiae* to express heterologous (fungal) genes with the utilization of enzyme assays. The fungus of interest is  
5 fermented under conditions that give high-level enzyme activity; mRNA is prepared from the resulting biomass and a cDNA library is constructed in *E. coli*. Plasmid DNA is isolated from subpools of this library and transformed into *S. cerevisiae*. Subsequently, the yeast transformants are screened for enzyme  
10 activity.

We assume that for a fungal genome about 5000 genes are expressed. For statistical reasons, and due to the manner in which cDNA is prepared, a high number of clones must be screened in order to ensure that all expressed enzymes are identified,  
15 i.e. on the order of magnitude of 50,000-100,000.

For a typical screen for any given enzyme or other gene product, a functional assay is applied: for example, proteases are screened in an assay specific for proteases, amylases are screened in an assay specific for amylases and so forth. The  
20 existing methods for traditional functional screening for extra-cellular enzymes are substantially limited to the applied screening assays. This means that screening of a genome provides a) only those enzymes for which a functional assay exists or can be designed, and b) only a single enzyme activity (or a very  
25 limited number of enzyme activities), i.e. the enzyme activity/activities that the assay is specific for or which can be derived from a single screening. Frequently, the same gene library is screened over and over again because it is desired to

investigate several activities. This is ideally done in parallel, but as it is often not known at the outset which enzymes are of potential interest, gene libraries have to be newly constructed from the given wild type organism or the library has to be screened several times in the various functional assays. This method for screening for enzymes or other proteins has the disadvantage of being both time-consuming and expensive.

An estimate of the total number of extracellular enzymes in *B. subtilis* was made by 2D gel analysis of extracellular enzymes, and subsequent identification of spots by N-terminal sequencing. The number was predicted to be 150-180 extracellular enzymes (Hirose et al. 2000. Proteome analysis of *Bacillus subtilis* extracellular Proteins: a two dimensional protein electrophoretic study. Microbiology 146:65-75). This means that with a screening procedure designed to identify all secreted gene products, the number of hits would be about 200 clones from the total of 4100 open reading frames. In other words, if 10,000 clones are screened, 200-500 clones will carry a DNA fragment from the original genome expressing an extra-cellular protein or peptide. With a pre-screening for clones producing extra-cellular enzymes functional screening, should, in the ideal situation, be able to be performed on these less than 500 clones.

For fungi, the number of secreted gene products is assumed to be in the range of about 500-1000 for a given genome, so that only about 500-1000 clones from a total of approximately 25,000-40,000 screened clones are of real interest.

A tremendous savings in both time and money could be achieved by mining the gene libraries or cDNA libraries for clones expressing extracellular products. In a typical example of screening of a bacterial genome, the gene library could thus  
5 be initially screened in a secretion assay in which 5000-10,000 clones are screened and the approximately 200 clones are detected that encode secreted gene products. These 200 clones could then be screened using e.g. functional assays.

This means that compared to a theoretical screening  
10 procedure based only on functional assays in which a gene library might be screened in 200 different functional assays to detect all secreted gene products (e.g. 5000 clones x 200 = 1,000,000 screened clones), in the ideal case in which clones producing secreted products may be initially identified, the  
15 gene library is screened once for secreted products (e.g. 5000 clones) and the resulting approximately 200 secreted clones can subsequently be fingerprinted for biochemical activity in functional assays. Assuming again use of the same 200 functional assays, a total of only 200 x 200 = 40,000 clones would have to  
20 be investigated, in other words only 4% of the 1,000,000 clones that would have to be investigated using a functional assay alone.

Typically, a gene library might be screened using about 10 different functional assays. With 5000 clones this gives a total  
25 of 50,000 clones that must be screened. In the ideal case in which clones producing secreted gene products can be identified at the outset, the 5000 clones are screened once, after which all secreting clones are detected and analysed in the 10

functional assays, corresponding to a screening workload of 200 clones screened in 10 assays, i.e. only 2000 clones need to be screened in the functional assays, again a total of only 4% of the number of clones that must be screened using the functional  
5 assays alone.

For fungal cDNA libraries, the same statistical considerations apply.

In short, it would be a tremendous advantage to have a screening assay for secreted enzymes and other proteins from  
10 gene libraries wherein the relatively few clones producing secreted gene products could be identified at the outset, so that only these few clones have to be investigated in functional assays aimed at identifying proteins of interest. The present invention provides such an assay.

15 Prior art includes disclosures such as WO 89/08114 and Analytical Biochemistry, Vol. 270 (1) pp. 103-111 (1999), which relates to the use of specific monoclonal antibodies to identify compounds such as expression product, e.g. surface bound antigenic products. The prior art techniques do however not  
20 pertain to solving the problem of identifying unknown compounds for which no suitable antibody can be identified.

#### **SUMMARY OF THE INVENTION**

The present invention provides a novel screening method that  
25 makes it possible to screen the genome of a microorganism only once in order to identify all gene products secreted by the microorganism. Clones producing these secreted gene products may then easily and quickly be further screened for peptides or

proteins having a desired function, e.g. enzymatic activity, and/or they may be subjected to nucleotide sequencing in order to identify genes encoding compounds of interest.

In a first aspect, the invention relates to a method for  
5 screening for compounds secreted by an organism, comprising

- (a) Raising antibodies against secreted products of a donor organism,
- (b) providing a gene library from the donor organism,
- (c) cloning the gene library into a suitable host organism,
- 10 (d) expressing the cloned genes in the host organism, and
- (e) detecting positive clones, expressing a cloned gene encoding a secreted compound, using the antibodies of (a) to identify such positive clones.

15 The invention also in a second aspect relates to a novel compound obtained or obtainable by the method of the invention.

In a third aspect the invention relates to a method for screening for a nucleotide sequence encoding a compound secreted by an organism comprising steps (a) to (e) of the first aspect  
20 and the additional step of subjecting at least one positive clone to nucleotide sequencing to identify at least one nucleotide sequence encoding a secreted compound.

In a fourth aspect the invention relates to a nucleotide sequence obtained or obtainable by the method of the third  
25 aspect.

In a fifth aspect the invention relates to a method for screening microorganisms for strains that secrete a compound comprising step (a) of the first aspect and the additional step

of subjecting the microorganism to an immunoassay using the antibodies from step (a) of the first aspect to identify microorganisms that secrete said compound.

In a sixth aspect the invention relates to a microorganism  
5 obtained or obtainable by the method of any of the fifth aspect.

#### DETAILED DESCRIPTION OF THE INVENTION

In the context of the present specification and claims, the term  
10 "secreted compound" refers to any compound that is secreted by a microorganism of interest. The compound may be associated with the membrane of the microorganism, but in particular a preferred secreted compound is a compound, which is secreted/transported by the microorganism beyond the cellular boundaries into the  
15 extracellular medium surrounding the cell entity. Secreted compounds include in particular, but are not limited to, proteins and peptides.

One object, which is solved by the present invention, is to find a method for identifying secreted compounds for which no  
20 functional assay exist and/or no antibody is known. Another object, which is solved by the present invention, is to find a method in which several, i.e. more than one, secreted component from a donor could be identified simultaneously, thus speeding up screening methods. A third object, which is solved by the  
25 present invention, is to find a method suitable for screening only secreted compounds.

As indicated above, secreted compounds such as enzymes are of particular interest. The enzymes screened for in accordance



with the invention may belong to known classes of enzymes, or they may be of unknown enzyme classes, e.g. enzymes having a desired functional activity but not necessarily belonging to a known enzyme class. As used herein, the term "enzyme class" (E.C.) refers to the internationally recognized enzyme classification system, Recommendations (1992) of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, Academic Press, Inc., 1992.

The types of enzymes which may appropriately be identified include oxidoreductases (EC 1.-.-.-), transferases (EC 2.-.-.-), hydrolases (EC 3.-.-.-), lyases (EC 4.-.-.-), isomerases (EC 5.-.-.-) and ligases (EC 6.-.-.-).

Preferred oxidoreductases in the context of the invention are peroxidases (EC 1.11.1), laccases (EC 1.10.3.2) and glucose oxidases (EC 1.1.3.4)], while preferred transferases are transferases in any of the following sub-classes:

- a) Transferases transferring one-carbon groups (EC 2.1);
- b) Transferases transferring aldehyde or ketone residues (EC 2.2); acyltransferases (EC 2.3);
- c) Glycosyltransferases (EC 2.4);
- d) Transferases transferring alkyl or aryl groups, other than methyl groups (EC 2.5); and
- e) Transferases transferring nitrogenous groups (EC 2.6).

A most preferred type of transferase in the context of the invention is a transglutaminase (protein-glutamine  $\gamma$ -glutamyltransferase; EC 2.3.2.13).

Preferred hydrolases in the context of the invention are: Carboxylic ester hydrolases (EC 3.1.1.-) such as lipases (EC 3.1.1.3); phytases (EC 3.1.3.-), e.g. 3-phytases (EC 3.1.3.8) and 6-phytases (EC 3.1.3.26); glycosidases (EC 3.2, which fall  
5 within a group denoted herein as "carbohydases"), such as  $\alpha$ -amylases (EC 3.2.1.1); peptidases (EC 3.4, also known as proteases); and other carbonyl hydrolases. Other preferred hydrolases are xyloglucanase, arabinase, rhamno-galactoronase, pectinases, ligninases (for example polyphenol hydrolase).

10 In the present context, the term "carbohydase" is used to denote not only enzymes capable of breaking down carbohydrate chains (e.g. starches) of especially five- and six-membered ring structures (i.e. glycosidases, EC 3.2), but also enzymes capable of isomerizing carbohydrates, e.g. six-membered ring structures  
15 such as D-glucose to five-membered ring structures such as D-fructose.

Carbohydases of relevance include the following (EC numbers in parentheses):

$\alpha$ -amylases (3.2.1.1),  $\beta$ -amylases (3.2.1.2), glucan 1,4- $\alpha$ -  
20 glucosidases (3.2.1.3), cellulases (3.2.1.4), endo-1,3(4)- $\beta$ -glucanases (3.2.1.6), endo-1,4- $\beta$ -xylanases (3.2.1.8), dextranases (3.2.1.11), chitinases (3.2.1.14), polygalacturonases (3.2.1.15), lysozymes (3.2.1.17),  $\beta$ -glucosidases (3.2.1.21),  $\alpha$ -galactosidases (3.2.1.22),  $\beta$ -galactosidases (3.2.1.23),  
25 mannanase (3.2.1.25), amylo-1,6-glucosidases (3.2.1.33), xylan 1,4- $\beta$ -xylosidases (3.2.1.37), glucan endo-1,3- $\beta$ -D-glucosidases (3.2.1.39),  $\alpha$ -dextrin endo-1,6- $\alpha$ -glucosidases (3.2.1.41),



The term "host organism" refers to any of the organisms into which a gene library of a donor strain can be displayed.

The term "gene library" includes both libraries of genomic nucleotide sequences, libraries of cDNA sequences and all other  
5 types of libraries of nucleotide sequences derived from a donor organism, e.g. a genomic DNA library obtained from a donor strain in accordance with the screening method of the invention. Preferably a gene library contains the entire genomic DNA or the entire cDNA from a donor organism.

10 The term "primary antibodies" refers to antibodies raised against a compound or a mixture of compounds secreted by the donor strain. The primary antibodies are preferably polyclonal and recognises the compound or compounds secreted by the donor organism. In a preferred embodiment the primary antibodies are  
15 raised against compounds present in the supernatant of a donor organism culture liquid and/or a broth from the fermentation of a donor organism. In the present screening method such primary antibodies will be capable recognising all compounds in the culture supernatant, i.e. secreted extracellular compounds. In  
20 another preferred embodiment the primary antibody or mixture of antibodies is raised against secreted extracellular compounds as well surface bound compounds. In the present screening method such antibody mixture will be capable recognising secreted compounds including secreted extracellular compounds and surface  
25 bound compounds being exposed to the extracellular medium. The primary antibodies may be labelled primary antibodies enabling detection of presence of a compound from the donor organism directly, or the primary antibodies may be recognized by

secondary antibodies. In a preferred embodiment the primary antibodies are a mixture of antibodies prepared by immunization of an animal with the culture supernatant of a donor organism or a culture supernatant including surface bound compounds of the donor organism and isolating the mixture of antibodies formed in the animal against secreted compounds of the culture supernatant.

The term "secondary antibodies" refers to labelled antibodies specifically recognizing the primary antibodies and enables detection of a bound primary antibody.

The terms "labelled primary antibodies" and "labelled antibodies" refer to antibodies labelled with a marker in order to detect their presence in a sample. This could be either by emission of radiation, by an enzyme reaction or by fluorescence.

The terms "heterologous expressed compound" and "heterologous expressed DNA" refer to a protein or other compound, respectively to genes of the donor organism expressed in a heterologous host organism that could be but are not restricted to *B. subtilis*, *E. coli*, *S. cerevisia* or *A oryza*.

The term "positive clone" refers to any host organism which (1) comprises a nucleotide sequence from a gene library derived from a donor organism under investigation, (2) upon cultivation produces and/or secretes a compound which gives an immunoreaction with primary antibodies, thereby enabling discrimination of said positive clone from other clones that also comprises a nucleotide sequence from a gene library derived from the donor organism, but do not produce and/or secretes a compound recognizable by primary antibodies.

The term "additional screening step" refers to any screening step performed in addition to the immunoassay screening of the invention. The additional screening step may in particular be a functional assay, for example an enzyme assay, 5 in which the function of said cloned gene is identified, or a receptor assay in which a secreted peptide binds to a receptor, or an antimicrobial assay in which an antimicrobial peptide interacts with a microorganism (bacterium or fungus).

According to the method of the invention, immunoassay 10 techniques are used to identify clones that produce a secreted gene product. A major advantage of this method is that even unknown gene products - for example an enzyme with no known biological function - can be detected. For such an enzyme, no functional screening assay will be available, but such enzymes 15 will be detected in a screening directed to secreted gene products.

The invention is a method for mining genomes of genes encoding all compounds secreted by a selected donor organism.

A culture supernatant is produced by cultivation of a donor 20 organism. Cells are separated from the supernatant. The supernatant, optionally including compounds bound or associated to the cell membrane, is used for immunization, so antibodies are raised against secreted compounds produced by the donor organism. The supernatant may be used directly for this 25 immunisation process or it may be refined prior to the immunisation process, e.g. by removing undesired compounds or adding desired compounds. However, as it is evident from objects of the present invention, that it is to be understood that

refinement of the supernatant do not include separating the secreted compounds into single discrete secreted compounds, and using these discrete secreted compounds to raise discrete antibodies, because such procedure would not provide the advantages of the present invention. Hence, it is essential to the invention, that the supernatant or mixture of secreted compounds used to raise antibodies comprise at least 2 secreted compounds of the donor organism, preferably more than 3, more preferably more than 5, more preferably more than 7, most preferably more than 10 secreted compounds of the donor organism to be of use in the present invention.

In another embodiment the mixture of antibodies may also be refined after recovering the mixture of antibodies from the immunisation process, of course with the above mentioned proviso that the refinement do not include separating the antibodies into single discrete antibodies. As an example of refinement of antibodies, the antibodies may be treated to remove or reduce antibodies which exhibits cross reaction with secreted compounds of a host organism, which in the present invention hosts the gene library of the donor organism, so as to reduce the background emanating from the host itself.

In the immunisation process antibodies are raised against all compounds in the refines or unrefined supernatant, optionally including compounds bound or associated to the cell membrane. Methods of raising of antibodies, e.g. polyklonal antibodies are known to the art such as from: N. M. G. Harboe & A. Ingild:

Immunization, Isolation of Immunoglobulins and Antibody Titre

Determination, Scand. J. Immunol., Vol. 17, Suppl. 10, pp 345-351, 1983.

A gene library comprising genomic DNA, preferably the entire genome of a donor organism or cDNA or mRNA, preferably  
5 the total mRNA of a donor organism may be constructed and inserted or cloned into a suitable host organism. The recovery of nucleotide material, insertion or cloning of the gene library into the host organism may be achieved by any suitable conventional method, such as transformation or transfection of  
10 the host organism with the library. Non-limiting examples of conventional methods for recovering nucleotide materials from a donor organism include the methods described in Pitcher et al., "Rapid extraction of bacterial genomic DNA with guanidium thiocyanate", Lett. Appl. Microbiol., 8, pp 151-156, 1989;  
15 Diderichsen et al., "Cloning of aldB, which encodes alpha-acetolactate decarboxylase, an exoenzyme from Bacillus brevis", J. Bacteriol., 172, pp 4315-4321, 1990; Dretzen et al., "A reliable method for the recovery of DNA fragments from agarose and acrylamide gels", Anal. Biochem., 112, pp 295-298, 1981 or WO 94/19454.  
20 Non-limiting examples of conventional methods for cloning nucleotide sequences into a suitable host organism includes the methods described in Ausubel, et al. (eds.) Current protocols in Molecular Biology, John Wiley and Sons, 1995, Harwood, C. R., and Cutting, S. M. (eds.), "Molecular Biological Methods for  
25 Bacillus", John Wiley and Sons, 1990 or Sambrook et al., "Molecular cloning: A laboratory manual", Cold Spring Harbor lab., Cold Spring Harbor, NY. (1989).



Although the primary objective of the invention is to screen a host organism comprising a gene library from a donor organism for clones producing a compound which is identical to a compound produced by the donor organism itself, the screening method of the invention may also be used to screen for clones producing a compound which is not completely identical to a compound produced by the donor organism itself, but which exhibits cross reactivity with primary antibodies. Cross reactivity between an antibody raised against compound produced by the donor organism and a non identical compound of a host organism requires a certain degree of similarity, but some differences in form of mutations and/or substitutions could be allowed, while preserving the feasibility of the present invention. Accordingly the gene library from the donor organism may be subjected to manipulation before cloning it into a suitable host cell e.g. by gene shuffling, such as described by Stemmer, Proc. Natl. Acad. Sci. USA, 91, pp. 10747-10751, 1994 and Stemmer, Nature, 370, pp. 389- 391, 1994; and/or by random mutagenesis as described by Eisenstadt et al., "Gene mutation", Methods for general and molecular bacteriology, pp. 297-316, Eds: Gerhardt P., Murray R.G.E., Wood W.A. and Krieg N.R., ASM, 1994.

Accordingly the invention encompasses screening methods wherein nucleotide sequences in the gene library of the donor organism have been mutated. In this context it is also possible by using the method of the invention to prepare gene libraries from other organisms than the donor organism, so as to identify secreted compounds from one or more other organisms which is



1) clones comprising the gene library are cultivated e.g. in micro plates.

2) supernatant or culture fluid is transferred to a hybridization membrane.

5 3) protein or other secreted compounds are immobilised on membrane or filter

4) primary antibodies are hybridized to the filter

5) detection of bound antibodies either with labelled secondary antibodies or directly with labelled primary antibodies.

10 Methods for dot blot hybridisation are known e.g. from Hawkes et al.; "A Dot Immuno Binding Assay for Mono Clonal and other Antibodies"; J. Analytical Biochemistry, Vol. 119 (1) pp. 142-147 (1982)

15 ELISA method a

1) clones comprising the gene library are cultivated e.g. in micro plates.

2) protein or other secreted compounds are immobilised on the micro plate

20 3) primary antibodies are hybridized to the filter

4) detection of bound antibodies either with labelled secondary antibodies or directly with labelled primary antibodies.

25 ELISA method b

1) clones comprising the gene library are cultivated e.g. in microplates.

2) primary antibodies are bound to another micro plate

- 3) supernatant or culture fluid with protein or other secreted compounds added to the plate coated with antibodies
- 4) detection of bound protein or alternative compound with labelled primary antibodies.

5 Methods for ELISA are known e.g. from Kirkegaard & Perry Laboratories. 1999 Product catalog. p. 66

#### Protein micro array:

- 1) clones comprising the gene library are cultivated e.g. in micro plates.
- 2) protein or other secreted compounds in the supernatant are immobilised on a microscope slide coated with Nylon or nitrocellulose
- 3) primary antibodies are hybridized to the filter
- 4) detection of bound antibodies either with labelled secondary antibodies or directly with labelled primary antibodies.

Methods for Protein micro array are known e.g. from Lueking et al.; "Protein microarrays for gene expression and antibody screening"; Analytical Biochemistry, Vol. 270 (1) pp. 103-111 (1999)

After screening the clones comprising the gene library, positive clones may be subjected to nucleotide sequencing using methods known per se in the art. The open reading frames of the corresponding inserts are determined and the corresponding gene identified. The gene may then be identified as falling into a known class of genes for which the functionality of the gene

products are known, or it may encode an unknown secreted product. Compared to the currently applied approach of whole genome sequencing, the method of the invention provides significant advantages in terms of speed and economy.

5 Methods for nucleotide sequencing are known in the art e.g. from Lee et al.; "New energy transfer dyes for DNA sequencing"; Nucleic Acids Research, 1997, vol. 25, No.14.

Further after the screening, the identified positive clones may be subjected to additional screening steps e.g. to verify  
10 the screening results and/or to achieve more specific differentiation between the positive clones. For verification the positive clones may be subjected to at least one additional screening comprising cultivating said positive clones and assaying them in a second immunoassay using the same antibodies  
15 as used in the first immunoassay to eliminate possible false positives.

Additional screening steps to achieve differentiation between the positive clones include subjecting the clones to functional assays, wherein functional properties of produced  
20 compounds are investigated. For enzyme compounds preferred functional properties includes but is not limited to wash performance, thermal stability, substrate specificity, catalytic turnover, oxidation stability, sensitivity to inhibitors, pH optimum, detergent stability, stability against microbial  
25 inactivation etc. For receptor binding peptides or peptide antibiotics preferred functional properties includes but is not limited to toxicology, distribution profile in the human or animal body, metabolism products, side effects, rate of

metabolisation or secretion, receptor binding capacity, antimicrobial capacity etc. Preferably additional screening steps are performed using as a starting material to be tested a supernatant obtained from cultivating positives clones or a refined and/or purified product thereof.

Enzymes and other proteins with unknown functions can prove to be of significant value in order to solve important needs. For example, in relation to enzymes used in laundry detergents, certain stains are known to be removable with known enzymes. However, it is desirable to be able to identify new enzymes or proteins that could be useful for removing such stains or other stains, although for this purpose the chemical nature of the stain is not necessarily known and the biochemical nature of the enzyme does not need to be known. By using a random approach based on identifying secreted enzymes or proteins in accordance with the invention, all secreted enzymes and other proteins of a given genome can be tested directly in a stain removal assay. This assay may for practical reasons not be applicable as a direct screening assay. Since the number of enzymes to be tested in this manner is very limited compared to the total number of gene products in the genome as a whole, the functional assay used to identify enzymes with a desired activity can be performed quickly and easily.

In relation to baking or other complex food processes enzymes may on the one hand catalyze reactions leading to useful products either in terms of taste, smell or texture. On the other hand enzymes may as well remove substrates that gives unwanted properties such as bad smell or odour or structure.

These reactions are not always obvious, and substrates and products are not necessarily known or may even be impossible to characterize. High throughput assays are not possible to do for such reactions. For this reason a pre-selection as disclosed  
5 herein is necessary.

In relation to animal feed enzymes may facilitate digestion and make feed better accessible or may even release useful compounds for the animal. Such assays may demand testing on animals. Also here a pre-selection is necessary in order to  
10 reduce the number of tested animals.

In relation to textile, it is well known that various hydrolytic enzymes are useful in the processing of especially cotton fabrics and yarn. The structure and the components of cotton are not fully known or understood, and for this reason  
15 enzymes of unknown function are useful. However, throughput in such tests is limited and again pre-selection is demanded.

A screening approach that detects secreted gene products in an initial step has a strong impact on the speed of the screening, the throughput needed, and the machinery and  
20 resources needed. Furthermore, this screening approach can also have a substantial impact on the nature of the encoded gene products to be identified. For example, a beneficial effect may be obtained where gene libraries from more than one genome are constructed and screened collectively, e.g. in environmental  
25 samples. It has been shown that a small sample of soil (e.g. 1 g) typically contains a variety of different microorganisms, containing both cultivable and so-called non-cultivable organisms (where the term "non-cultivable" refers to organisms

that cannot be cultivated in any known medium) corresponding to 4000 different genomes (Torsvik et al. 1990. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria, Appl. Environ. Microbiol., 56: pp 776-781).

5 Typically, several hundred thousand clones are screened in a given functional assay to detect clones carrying a desired activity. Again, by using a secretion assay according to the invention, organisms, which is different from a donor organism from which secreted compounds antibodies are raised, which  
10 secretes similar compounds can be identified and in case of screening of gene libraries from such organisms a significantly lower number of clones need to be screened, which can have a drastic effect on the throughput and workload.

The above-described system could also be applied to a  
15 consortium of organisms subjected to a multi-resistant pathogenic bacterium. Under such conditions, antimicrobial agents would be induced. Subsequently, the method could selectively mine the library of the entire consortium for clones expressing antimicrobial agents.

20 The above-mentioned system could be applied for a consortium of organisms subjected to a bacterium that might be multi resistant to known antibiotic and might be pathogenic to humans animals or plants. Under such conditions antimicrobial agents would be induced. Subsequently the method could  
25 selectively mine the library of the entire consortium for clones expressing antimicrobial agents. It is impossible to perform such tests with the extremely high throughput of a normal



procedure. However, it will be possible by using the antibody mining disclosed here.

With cell cultures the method can be applied for mining for potential receptor binding compounds.

5 Summarizing, the present invention provides a new approach in which e.g. a bacterial or fungal genome is screened only once in order to detect all secreted gene products. These secreted gene products can then be subjected to additional assays, such as specific functional screening assays, to identify clones of  
10 interest and for e.g. identifying novel compounds with a desired activity.

The invention will be further illustrated in the following non-limiting examples.

## 15 **EXAMPLES**

### **Example 1:**

Screening for multiple proteins by colony hybridisation

20 Production of antibodies

Polyclonal antibodies against extra-cellular enzymes from a bacterium, *Bacillus agaradhaerens* strain AC13, was made by cultivation of the strain in BPX medium supplemented with 100 mM  
25 NaHCO<sub>3</sub> at 30°C for 5 days with agitation at 250 RPM (BPX medium is: potato flour 100 g/l, barley flour 50 g/l, sodium caseinate 10 g/l, soy-bean cake 20 g/l, Na<sub>2</sub>HPO<sub>4</sub>·12H<sub>2</sub>O, 9 g/l and Pluronic 0.1 ml/l autoclaved in 100 ml aliquots for 40 min at 120°C). The

supernatant of 50 ml of the culture was recovered by centrifugation at 3000 RPM, and concentrated by freeze drying. Subsequently, the freeze dried protein was re-suspended in 10 ml of water, sterile filtered and used for raising antibodies in 5 rabbit (The immunisation, and recovery of antibodies provided by DAKO, Copenhagen, Denmark). These antibodies was referred to as "primary antibodies".

#### 10 Genomic DNA Preparation

Bacillus agaradhaerens strain AC13 was propagated in LB medium to which 100 mM NaHCO<sub>3</sub> had been added. After 16 hours of incubation at 30°C and 300 RPM, the cells were harvested, and genomic DNA was isolated by the method described by Pitcher et al. (Pitcher, D. G., Saunders, N. A., Owen, R. J . 1989. Rapid 15 extraction of bacterial genomic DNA with guanidium thiocyanate; Lett. Appl. Microbiol. 8:151-156). (LB was: 25 g/l of LB bouillon Merck Art. 0285; for agar 15 g/l of Bacto-agar Difco Art. 0140 was added prior to autoclaving).

#### 20 Genomic Library Construction

Genomic DNA was partially digested with restriction enzyme Sau3A and size-fractionated by electrophoresis on a 0.7% agarose gel. Fragments of between 2 and 7 kB in size were isolated by 25 electrophoresis onto DEAE-cellulose (Dretzen G, Bellard M, Sassone-Corsi P, Chambon P. 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels;

Anal. Biochem. 112 295-298). Isolated DNA fragments were ligated to BamHI digested pSJ1678 plasmid DNA.

Ligated DNA was used in electroporation of *E. coli* SJ2, or *B. subtilis* PL 3350. The transformed cells were plated onto LB-agar plates containing 10 ppm (for *E. coli*) or 6 ppm (for *B. subtilis*) chloramphenicol, and the agar plates were incubated for 18 hours at 37°C.

#### 10 Colony hybridisation

The resulting colonies were blotted to a nitrocellulose filter (Schleicher & Schuell Art. 10401116) by incubating the agar plate with the filter on top for 2 h at room temperature. The filter was incubated for 10 min in a blocking solution, 15 incubated in washing buffer pH 10 to which 1000x diluted primary antibodies was added, washed three times in washing buffer, incubated in washing buffer to which 1000x diluted anti-rabbit antibody (secondary antibody, Art. No. P0217, DAKO, Copenhagen) conjugated with horse radish peroxidase (HRP) had been added, 20 washed three times in washing buffer, washed once in acetate buffer pH 5.5, and incubated in staining solution to develop colour. All these steps were carried out at room temperature and with agitation, 50 RPM. Positive clones were detected by the formation of red spots on the filter.

25

Solutions for colony hybridization

Washing buffer: TRIS Merck Art. 108382, 6 g/l; Sodium chloride Merck Art. 106404, 8.75 g/l and Tween 20 Merck Art. 822184, 0.5 g/l in deionised water, pH 10

5 Blocking solution: Tween 20 Merck Art. 822184, 20 g/l in washing buffer

10 Acetate buffer pH 5.5: Sodium acetate Merck Art. 106267, 6.8 g/l, and acetic acid Merck art. 100063, 1.8 ml/l in deionised water.

Staining solution: 3-amino-9-ethylcarbazole stock 40 ml/l, Hydrogen peroxide Merck Art. 7209, 0.5 ml/l.

15 3-amino-9-ethylcarbazole stock: 5 g/l of 3-amino-9-ethylcarbazole Sigma Art. A-5754 in absolute ethanol.

Positive clones were recovered from the agar plate and fermented in LB broth containing 6 ppm chloramphenicol for 2 days at 37°C  
20 with agitation at 250 RPM. Cells were removed by centrifugation, 5 ml supernatant was loaded onto a 10% Bis-Tris gel (Novex Art. P0217). The gel was run as recommended by the manufacturer. The gel was then blotted to a nitrocellulose filter (BA85 Protran, Schleicher & Schuell) using a semi-dry blotter (Semi dry blotter  
25 II, Kem-En-Tec, Denmark). The filter was hybridised and stained as described above. A positive hit was seen as a discrete band on the filter.

### **Example 2:**

Screening for multiple proteins with dot blot hybridisation of supernatants in micro-plates

5 The DNA library of Example 1 was plated out and cultivated as described above. Colonies were picked from plates and transferred to micro-plates using an automated colony picker (Flexys colony picking station, Genomic Solutions Inc., Ann Arbor). The DNA library was grown in micro-plates as described  
10 for liquid fermentation above.

The plates were centrifuged and supernatants were transferred to a nitrocellulose filter. The filter was hybridised and stained as described above. Samples of 1  $\mu$ l of culture supernatants from  
15 cultures grown in TY broth for 1 to 3 days were loaded to the filter. Red spots on the filter indicated a reaction with the antibodies.

### **Example 3:**

20 Screening for multiple proteins with protein arrays

The DNA library of Example 1 was plated out and cultivated as described above. Colonies were picked from the plates and transferred to micro-plates using an automated colony picker  
25 (Flexys colony picking station, Genomic Solutions Inc, Ann Arbor). The DNA library was grown in micro-plates as described for liquid fermentation above. These plates were then centrifuged at 3000 RPM and supernatants were transferred to new

micro-plates. The micro-plate supernatants were spotted onto a nitrocellulose-coated microscope slide (Oncyte Art. No. 70332, Electron Microscopy Sciences, Fort Washington, Pennsylvania) using an arrayer (GMS417, Affymetrics, Santa Clara, California).

5 The array was hybridised as described above, except that the secondary antibody was anti-rabbit antibody labelled with Alexafluor-546 (Art. No. A-11035, Molecular Probes) The array was read using an array scanner (GMS418, Affymetrics, Santa Clara, California). The sensitivity of this method was improved  
10 by making a normalisation to the total amount of protein loaded at the array using Sypro orange (Molecular Probes Art. S-6650) staining. The array was incubated for 5 minutes in Sypro orange diluted 5000x in washing buffer (see above) followed by a brief wash in washing buffer. Protein was measured as red fluorescence  
15 with the array scanner. Normalisation and identification of positive hits were made with the computer programs Genesight, Image, and Clonetracker (Affymetrics, Santa Clara, California).

#### 20 **Example 4**

Screening for multiple fungal proteins with protein arrays

##### *Production of antibodies*

Polyclonal antibodies against extra-cellular compounds from a  
25 fermented fungus, *Humicola insolens*, was made by (1) recovery of a *Humicola insolens* fermentation broth and raising antibodies in rabbit against the supernatant of this fermentation broth. The procedure was as described in example 1.

1 The antibody mixture raised against the extra-cellular compounds  
2 secreted from the *Humicola insolens* fermentation broth was  
3 evaluated by Western Blot analysis of relevant fermentation  
4 samples. This analysis revealed around 20 bands in the western  
5 blot demonstrating a broad spectrum of antibodies.

#### *Pre-treatment of antibodies*

6 The prepared antibody mixture was subjected to pre-treatment  
7 with the supernatant of an *A. oryzae* transformation host  
8 fermentation to reduce or eliminate background from  
9 transformation host. This background was probably caused by  
10 unspecific antibody response from glycosylated host proteins.  
11 The pre-treatment was carried out by mixing 100 µl antibody  
12 mixture with 400 µl *A. oryzae* supernatant and 5 ml tris malein  
13 buffer pH 7.0. This mixed solution was shaken for 2 hours at  
14 room temperature. After centrifugation at 4000 rpm the  
15 supernatant was mixed ad 100 ml with washing buffer pH 10,  
16 described in example 1.

17 The Tris/malein buffer, pH 7.0, was prepared by dissolving 1.2  
18 g/l Malein acid and 2.6 g/l Tris 7-9 in water.

#### *cDNA library construction*

19 A EcoR1/NotI directional *Humicola insolens* cDNA library was  
20 constructed in a AMA1-containing *Aspergillus* expression vector  
21 based on the autonomously replicating element AMA as described  
22 in WO 00/24883 A1 (see page 2, line 17-26 and the examples on  
23 page 36, line 7 to page 40, line 3, incorporated herein by  
24 reference). The *H. insolens* poly(A)+RNA was isolated from a *H.*

*insolens* production fermentation, 33 hours after fermentation start.

Double-stranded cDNA was synthesized from 5 µg of *Humicola insolens* poly(A)<sup>+</sup> RNA by the RNase H method (Gubler and Hoffman 5 1983, *supra*; Sambrook et al., 1989) using a hair-pin modification. The poly(A)<sup>+</sup>RNA (5 µg in 5 µl of DEPC-treated water) was heated at 70°C for 8 minutes in a pre-siliconized, RNase-free Eppendorf tube, quenched on ice, and combined in a final volume of 50 µl with reverse transcriptase buffer (50 mM 10 Tris-Cl pH 8.3, 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM DTT) containing 1 mM of dATP, dGTP and dTTP, and 0.5 mM of 5-methyl-dCTP, 40 units of human placental ribonuclease inhibitor, 4.81 µg of oligo(dT)<sub>18</sub>-NotI primer (Amersham-Pharmacia Biotech, Uppsala, Sweden) and 1000 units of SuperScript II reverse transcriptase (Gibco-BRL, 15 USA).

First-strand cDNA was synthesized by incubating the reaction mixture at 45°C for 1 hour. After synthesis, the mRNA:cDNA hybrid mixture was gel filtrated through a Pharmacia MicroSpin S-400 HR spin column according to the manufacturer's 20 instructions.

After the gel filtration, the hybrids were diluted in 250 µl of second strand buffer (20 mM Tris-Cl pH 7.4, 90 mM KCl, 4.6 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.16 mM βNAD<sup>+</sup>) containing 200 µM of each dNTP, 60 units of *E. coli* DNA polymerase I (Pharmacia, 25 Uppsala, Sweden), 5.25 units of RNase H, and 15 units of *E. coli* DNA ligase. Second strand cDNA synthesis was performed by incubating the reaction tube at 16°C for 2 hours, and an additional 15 minutes at 25°C. The reaction was stopped by



addition of EDTA to 20 mM final concentration followed by phenol and chloroform extractions.

The double-stranded cDNA was ethanol precipitated at -20°C for 12 hours by addition of 2 volumes of 96% ethanol and 0.2 volume of 10 M ammonium acetate, recovered by centrifugation, washed in 70% ethanol, dried (SpeedVac), and resuspended in 30 µl of Mung bean nuclease buffer (30 mM sodium acetate pH 4.6, 300 mM NaCl, 1 mM ZnSO<sub>4</sub>, 0.35 mM dithiothreitol, 2% glycerol) containing 25 units of Mung bean nuclease. The single-stranded hair-pin DNA was clipped by incubating the reaction at 30°C for 30 minutes, followed by addition of 70 µl of 10 mM Tris-Cl, pH 7.5, 1 mM EDTA, phenol extraction, and ethanol precipitation with 2 volumes of 96% ethanol and 0.1 volume 3 M sodium acetate pH 5.2 on ice for 30 minutes.

The double-stranded cDNAs were recovered by centrifugation (20,000 rpm, 30 minutes), and blunt-ended with T4 DNA polymerase in 30 µl of T4 DNA polymerase buffer (20 mM Tris-acetate, pH 7.9, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM dithiothreitol) containing 0.5 mM of each dNTP, and 5 units of T4 DNA polymerase by incubating the reaction mixture at +16°C for 1 hour. The reaction was stopped by addition of EDTA to 20 mM final concentration, followed by phenol and chloroform extractions and ethanol precipitation for 12 h at -20°C by adding 2 volumes of 96% ethanol and 0.1 volume of 3M sodium acetate pH 5.2.

After the fill-in reaction the cDNAs were recovered by centrifugation as above, washed in 70% ethanol, and the DNA pellet was dried in a SpeedVac. The cDNA pellet was resuspended

10  
15  
20  
25

in 25 µl of ligation buffer (30 mM Tris-Cl, pH 7.8, 10 mM MgCl<sub>2</sub>, 10 mM dithiothreitol, 0.5 mM ATP) containing 2 µg *EcoRI* adaptors (0.2µg/µl, Pharmacia, Uppsala, Sweden) and 20 units of T4 ligase by incubating the reaction mix at 16°C for 12 hours. The  
5 reaction was stopped by heating at 65°C for 20 minutes, and then placed on ice for 5 minutes. The adapted cDNA was digested with *NotI* by addition of 20 µl autoclaved water, 5 µl of 10x *NotI* restriction enzyme buffer and 50 units of *NotI*, followed by incubation for 3 hours at 37°C. Heating the sample at 65 °C for  
10 15 minutes stopped the reaction. The cDNAs were size-fractionated by agarose gel electrophoresis on a 0.8% SeaPlaque GTG low melting temperature agarose gel (FMC, Rockland, ME) in 1x TBE (in autoclaved water) to separate unligated adaptors and small cDNAs. The gel was run for 12 hours at 15 V, and the cDNA  
15 was size-selected with a cut-off at 0.7 kb by cutting out the lower part of the agarose gel. Then a 1.5% agarose gel was poured in front of the cDNA-containing gel, and the double-stranded cDNAs were concentrated by running the gel backwards until it appeared as a compressed band on the gel. The cDNA-  
20 containing gel piece was cut out from the gel and the cDNA was extracted from the gel using the GFX gel band purification kit (Amersham, Arlington Heights, IL) as follows. The trimmed gel slice was weighed in a 2 ml Biopure Eppendorf tube, then 10 ml of Capture Buffer was added for each 10 mg of gel slice, the gel  
25 slice was dissolved by incubation at 60°C for 10 minutes, until the agarose was completely solubilized, the sample at the bottom of the tube by brief centrifugation. The melted sample was transferred to the GFX spin column placed in a collection tube,

incubated at 25°C for 1 minute, and then spun at full speed in a microcentrifuge for 30 seconds. The flow-through was discarded, and the column was washed with 500 µl of wash buffer, followed by centrifugation at full speed for 30 seconds. The collection tube was discarded, and the column was placed in a 1.5 ml Eppendorf tube, followed by elution of the cDNA by addition of 50 µl of TE pH 7.5 to the center of the column, incubation at 25°C for 1 minute, and finally by centrifugation for 1 minute at maximum speed. The eluted cDNA was stored at -20°C until library construction.

Ligated cDNA was used in electroporation of *E. coli* Electromax DH10B, and DNA amplification carried out. Library pools of approx. 20000 cfu were plated out, and Qiagen-purified pool DNA thereof was used for transformation in the *A. oryzae* host.

Spore pools were harvested from Cove agar plates, and 29 pools was prepared, each pool representing 3 agar plates containing approx. 300 colonies.

#### *Screening of cDNA library by using protein arrays*

The cDNA library was inoculated and grown in 96 well low-protein-binding filtration plates (Millipore, MAHVS45). The primary 96-well plate screens involved the dilution of spores from distinct mutant pools into fermentation substrate so that one spore in average was inoculated per well when 100 µl of medium was dispensed into each well. After inoculation, the plates were incubated for 3-4 days at 30-34°C under static conditions in a culture box as described in WO 01/32844. A

relatively concentrated micro titer plate fermentation substrate, was used to ensure high sensitivity.

The micro titer plate fermentation substrate contained:

MgSO<sub>4</sub>, 7H<sub>2</sub>O : 0.3 g/l  
5 K<sub>2</sub>SO<sub>4</sub> : 0.3 g/l  
KH<sub>2</sub>PO<sub>4</sub> : 5.0 g/l  
tracemetal 1 : 0.25 ml/l  
urea : 0.65 g/l  
yeast extract : 0.5 g/l

10 in demineralized water and was pH adjusted to 6.5 with NaOH.

After sterilization 5 g/l carbon-source (5000 mg/l maltose) was added.

The plates were then vacuum filtered into new micro-plates. The filtrated samples were spotted onto a nitrocellulose-coated  
15 microscope slide (Oncyte Art. No. 70332, Electron Microscopy Sciences, Fort Washington, Pennsylvania) using an arrayer (GMS417, Affymetrics, Santa Clara, California). The array was hybridised as described above in example 3, except that the pretreated antibody mixture was used. Also reduced antibody  
20 concentration in the hybridization step was used to further decrease background.

The array was read using an array scanner (GMS418, Affymetrics, Santa Clara, California). Identification of positive hits was made with the computer programs Genesight, Imogene, and  
25 Clonetracker (Affymetrics, Santa Clara, California) and positive hits was isolated on CoveN agar and retested.

For isolated transformants of interest single colonies was transferred to agar slants. Spores from agar slants were

inoculated into 96-well plates with approx.  $10^3$  spores per well and fermented under static conditions for 4 days at 34°C to retest isolates. Subsequently, selected isolates were fermented in shake flasks.

5 In the screening of the *Humicola insolens* cDNA library 4500 clones were screened, and of these 166 hits was recorded or identified. In the screening method positive control strains were confirmed as positive, and host strain controls were confirmed negative. Retesting including protein array with  
10 varying conditions, SDS-Page and western blot was carried out. Retest in microtiter plates by using protein arrays confirmed 83 % of the isolates as positive. In SDS-PAGE testing 42 out of 44 clones randomly selected from the 166 hits had a significant, high expression of a heterologous protein as judged by SDS-PAGE,  
15 representing >8 different molecular weights.

In western blot testing 32 out of 33 clones randomly selected from the 166 hits produced a significant band verifying these clones as hits.

Eighteen clones, representing at least eight different  
20 molecular weights was selected for sequencing and sequence data confirmed the variety of clones.

This example confirms the versatility and speed of the present invention in screening for unknown secreted compounds of a donor organism. Further, in the screening for new enzymes this  
25 method may be simultaneously combined with enzyme activity measurements so that an even further classification of the hits can be achieved.